

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-101919  
 (43)Date of publication of application : 07.04.2000

(51)Int.CI. H04N 5/278  
 H04N 5/225

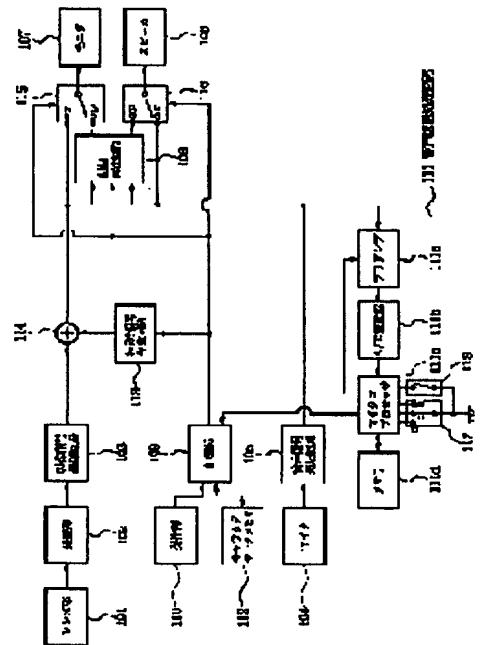
(21)Application number : 10-273129 (71)Applicant : CANON INC  
 (22)Date of filing : 28.09.1998 (72)Inventor : TAKEI HIROFUMI

## (54) SIGNAL PROCESSOR AND METHOD, REPRODUCING DEVICE AND METHOD, AND IMAGE PICKUP DEVICE

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To easily display plural characters and to display a character title on a screen with more effective timing.

**SOLUTION:** This signal processor processes both voice and image signals and consists of an input means which inputs the voice signals, a voice recognition means 111 which has a voice registration mode to register the input voices, in response to a prescribed character and a voice recognition mode to recognize a prescribed voice signal among those input voices and recognizes the inputted voice signals, a generation means 113 which generates character signals corresponding to the inputted voice signals, a display means which displays characters corresponding to the character signals produced by the means 113 and a recording means which records the character signals produced by the means 113 together with the image signals.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) 1998,2003 Japan Patent Office

## \* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

## CLAIMS

## [Claim(s)]

[Claim 1] An input means to be equipment which processes a sound signal and a video signal, and to input a sound signal. A speech recognition means to have the speech recognition mode in which the predetermined sound signal in the voice register mode which is made to correspond to a predetermined character and registers said input voice, and said input voice is recognized, and to recognize said inputted sound signal. A generating means to generate the character signal corresponding to said recognized sound signal. The signal processor characterized by having a display means to display the character corresponding to the character signal generated by said generating means, and a record means to record the character signal generated by said generating means with a video signal.

[Claim 2] Said speech recognition means is a signal processor according to claim 1 characterized by recognizing the voice registered beforehand from the inside of said input voice.

[Claim 3] Said speech recognition means is a signal processor claim 1 characterized by having the register mode which is made to correspond to a character and registers said input voice, or given in two.

[Claim 4] Said speech recognition means is a signal processor according to claim 3 characterized by matching and memorizing the control information which has memory and starts said character signal to said memory, and said input voice.

[Claim 5] It is a signal processor claim 3 characterized by matching with said character and registering said input voice while said speech recognition means indicates by the character with said display means in said register mode, or given in four.

[Claim 6] It is the signal processor according to claim 5 characterized by matching with said two or more characters, and registering said input voice while said speech recognition means displays said two or more characters with said display means in said register mode.

[Claim 7] It is the signal processor according to claim 6 which is equipped with a selection means to choose two or more characters displayed by said display means at the time of said register mode, and is characterized by said speech recognition means registering said input voice inputted to the character chosen by said selection means.

[Claim 8] Said speech recognition means is a signal processor according to claim 3 to 7 characterized by judging whether said input sound signal is suitable for registration in said register mode, and registering said input voice according to the judgment result concerned.

[Claim 9] It is the signal processor according to claim 8 characterized by controlling said character generating means and said display means in order to perform a predetermined display while forbidding registration of the input sound signal concerned, when it judges with said input sound signal of said speech recognition means being unsuitable to registration.

[Claim 10] Said speech recognition means is a signal processor according to claim 1 to 9 characterized by outputting the control information which starts said character based on said recognized sound signal to said generating means.

[Claim 11] Said speech recognition means is a signal processor according to claim 10 which has the memory which memorizes said control information, reads said control information from said memory based on said recognized sound signal, and is outputted to said generating means.

[Claim 12] It is the signal processor according to claim 11 characterized by matching said memory further and memorizing said control information and speech information, for said speech recognition means reading the control information corresponding to said recognized input sound signal from said memory while recognizing said input sound signal using said speech information, and outputting to said generating means.

[Claim 13] Said speech recognition means is a signal processor according to claim 12 characterized by detecting the degree of correlation with the speech information memorized by said memory and said input sound signal, and recognizing said input sound signal based on the detection result concerned.

[Claim 14] Said speech recognition means is a signal processor according to claim 13 characterized by controlling said character generating means and said display means in order to perform a predetermined display while forbidding outputting said control information from said memory to said generating means, when the degree of said correlation is low.

[Claim 15] Said generating means is a signal processor according to claim 10 to 14 characterized by reading the character data according to the control information which was equipped with a storage means to memorize the character data corresponding to two or more characters, respectively, and was outputted from said speech recognition means from said storage means.

[Claim 16] It is the signal processor according to claim 1 to 15 which picturizes a photographic subject, is equipped with an image pick-up means to output a video signal, and is characterized by carrying out multiplex [ of said display means ] to the image corresponding to the video signal outputted from said image pick-up means, and displaying the character corresponding to the character signal from said generating means on it.

[Claim 17] Said speech recognition means is a signal processor according to claim 1 characterized by having a manual setting-out means to set up said voice register mode and said speech recognition mode.

[Claim 18] Claim 1 characterized by having the voice input switch which turns into a trigger switch for beginning registration actuation and recognition actuation in said voice register mode and said speech recognition mode, or a signal processor given in 17.

[Claim 19] Said generating means is a signal processor according to claim 1 to 18 characterized by suspending generating of said character signal after predetermined time amount after starting generating of said character signal.

[Claim 20] A regenerative apparatus equipped with a playback means to be equipment which reproduces a sound signal and a video signal, and to reproduce a sound signal, a speech recognition means to recognize said reproduced sound signal, a generating means to generate the character signal corresponding to said recognized sound signal, and a display means display the character corresponding to the character signal generated by said generating means.

[Claim 21] Said speech recognition means is a regenerative apparatus according to claim 20 characterized by having the voice register mode which is made to correspond to a predetermined character and registers input voice, and the speech recognition mode in which the predetermined sound signal in said playback voice is recognized.

[Claim 22] Said speech recognition means is a signal processor according to claim 21 characterized by having a manual setting-out means to set up said voice register mode and said speech recognition mode.

[Claim 23] Claim 21 characterized by having the voice input switch which turns into a trigger switch for beginning recognition actuation in said voice register mode, or a regenerative apparatus given in 22.

[Claim 24] Said speech recognition means is a regenerative apparatus according to claim 20 to 23 characterized by judging whether playback sound signal level is more than predetermined level, and performing recognition actuation according to the judgment result concerned.

[Claim 25] Said speech recognition means is a regenerative apparatus according to claim 24 characterized by starting recognition actuation when it judges with said playback voice level being more than predetermined level.

[Claim 26] Input a sound signal, make it correspond to a character, and said input sound signal is registered. While displaying the character corresponding to the character signal which has

recognized said inputted sound signal based on said registered sound signal, generated the character signal corresponding to said recognized sound signal, and was generated by said generating means. The signal-processing approach characterized by recording the character signal generated by said generating means with a video signal.

[Claim 27] The playback approach characterized by displaying the character corresponding to the character signal which reproduced the sound signal, has recognized said reproduced sound signal, generated the character signal corresponding to said recognized sound signal, and was generated by said generating means.

---

[Translation done.]

**\* NOTICES \***

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

**DETAILED DESCRIPTION****[Detailed Description of the Invention]****[0001]**

**[Field of the Invention]** Especially this invention relates to the equipment and the approach accompanied by a character display about a signal processor and an approach, a regenerative apparatus, an approach, and image pick-up equipment.

**[0002]**

**[Description of the Prior Art]** Video camera one apparatus VTR is known as equipment which processes conventionally the video signal photoed with the video camera. And in video camera one apparatus VTR, it had the character display circuit and there was a thing equipped with the function which compounds and displays a character title on the photoed video signal.

**[0003]** In this kind of equipment, when a character title was displayed on a screen, the photography person was performing actuation by the switch.

**[0004]**

**[Problem(s) to be Solved by the Invention]** However, not only actuation is complicated, but by such approach, the display timing of a character title is not in agreement with a motion of a screen for actuation by hand control. Therefore, it was difficult to display a character title with sufficient timing on a screen. It was much more difficult especially to display two or more character titles according to the content of a screen.

**[0005]** This invention aims at solving the problem like the above-mentioned.

**[0006]** The object of further others of this invention is in the place which enables two or more character displays easily, and enables the display of a character title on a screen to more effective timing.

**[0007]**

**[Means for Solving the Problem]** In the signal processor according [ on the bottom of such an object, and ] to invention given in this claim 1 It has the speech recognition mode in which the predetermined sound signal in an input means to input a sound signal, the voice register mode which is made to correspond to a predetermined character and registers said input voice, and said input voice is recognized. It considered as the configuration equipped with a speech recognition means to recognize said inputted sound signal, a display means to display the character signal corresponding to said recognized sound signal on a screen, and a record means to record said displayed character signal with a video signal.

**[0008]**

**[Embodiment of the Invention]** Hereafter, the operation gestalt of this invention is explained to a detail using a drawing.

**[0009]** Drawing\_1 is the block diagram showing the configuration of the camera one apparatus digital VTR with which this invention is applied as first operation gestalt.

**[0010]** First, the actuation at the time of the usual record playback is explained. In drawing, the lens section 101 leads photographic subject light to the image pick-up section 102 including a lens, a zoom device, an iris, etc. The image pick-up section 102 picturizes a photographic subject using CCD, changes the obtained photographic subject light into an electrical signal, and outputs it to the video-signal processing circuit 103.

[0011] To the signal outputted from the image pick-up section 102, a clamp, gamma amendment, a white balance, etc. process common knowledge for making the picturized signal into the video signal based on a television signal, and the video-signal processing circuit 103 outputs them to the record regenerative circuit 106 and a switch 115 through an adder 114. Here, like the after-mentioned, the character signal outputted from the character signal creation circuit 113 is added to the video signal outputted from the video-signal processing circuit 103 in the adder 114, and an adder 114 outputs it to the record regenerative circuit 106 and a switch 115.

[0012] The record regenerative circuit 106 performs processing required in order to compress the amount of information and to record [-ization / addition of a synchronization, ID, etc., / error correcting code] it to the compressed video signal by encoding using well-known variable length coding, such as A/D conversion, DCT, and Huffman coding, to the video signal outputted from the video-signal processing circuit 103, and changes it into the gestalt suitable for record.

Furthermore, the video signal which performed such processing is recorded on a magnetic tape. It connects with r side edge child at the time of video-signal record, and a switch 115 outputs the video signal outputted from the video-signal processing circuit 103 to a monitor 107.

[0013] Thus, the recorded video signal is outputted to a switch 115, after the record regenerative circuit 106 is reproduced and an error correction, variable-length decode, Reverse DCT, etc. are processed. It connects with p side edge child at the time of video-signal playback, and a switch 115 outputs the regenerative signal from the record regenerative circuit 106 to a monitor 107.

[0014] On the other hand, a microphone 104 collects voice and outputs the obtained voice to the sound signal processing circuit 105. To the sound signal from a microphone, level adjustment, noise reduction, etc. process common knowledge, and output the sound signal processing circuit 105 to the record regenerative circuit 106, a switch 116, and the speech recognition circuit 111. To the sound signal from the sound signal processing circuit 105, A/D conversion, shuffling, an error correction, etc. process common knowledge, and record the record regenerative circuit 106 on a magnetic tape. It connects with r side edge child at the time of sound signal record, and a switch 116 outputs the sound signal outputted from the sound signal processing circuit 105 to a loudspeaker 108. About the speech recognition circuit 111, it mentions later.

[0015] Thus, the recorded sound signal is outputted to a switch 116, after the record regenerative circuit 106 is reproduced and an error correction, a DESHAFU ring, etc. are processed. It connects with p side edge child at the time of sound signal playback, and a switch 116 outputs the regenerative signal from the record regenerative circuit 106 to a loudspeaker 108.

[0016] Next, actuation of the speech recognition circuit 111 in this gestalt is explained. Generally there is an object for unspecified speakers which does not limit the object for specified speakers and speaker who limit a speaker, but is recognized also with whose voice in a voice recognition unit. The object for specified speakers makes the vocabulary to recognize utter beforehand, registers it, and recognizes based on this registered data. Since a recognition system can be assigned to the specific speaker who uses it according to the equipment for these specified speakers, while the load of a system becomes light, a high recognition rate can be expected, and it is hard to be dependent also on language.

[0017] On the other hand, since the object for unspecified speakers has unnecessary registration actuation, there is simple nature which a speaker is not chosen [ nature ] but operates speech recognition immediately, but in order to raise recognition precision, an arithmetic unit and a system with large-scale memory are needed. Since that a user is limited to about 1 individual in the case of the camcorder/movie like this gestalt, small, and low cost are absolutely made into conditions, it is a specified speaker and it can be said that the voice recognition unit for a specific vocabulary is further suitable.

[0018] In this gestalt, the voice recognition system corresponding to a specified speaker is used from such a background. For this reason, the user was made to utter the vocabulary to recognize beforehand and it has the voice register mode which registers that voice data as a reference voice pattern, and the speech recognition mode in which matching with input voice and a reference voice pattern is performed.

[0019] In drawing 1, a switch 117 performs a change-over in voice mode, and the switch 117 is selectable in the position c which turns off a voice input function, voice register mode a, and three positions in the speech recognition mode b. A switch 118 is a voice input switch used as the trigger switch at the time of inputting registration and the voice which a user utters at the time of recognition mode.

[0020] Next, a character signal is generated by the speech recognition result, and the case where it displays on a monitor 107 is described. The speech recognition circuit 111 where speech recognition processing is performed to the sound signal inputted into the speech recognition circuit 111 through the sound signal processing circuit 105 from the microphone 104 in the speech recognition circuit 111 like the after-mentioned outputs the speech recognition result to a control section 109. The speech recognition circuit 111 operates here, only when set as voice register mode or speech recognition mode by the switch 117.

[0021] A control section 109 reads the character data according to the speech recognition result from the speech recognition circuit 111 from the character data memory 112, and outputs it to the character signal creation circuit 113. The character signal creation circuit 113 changes the character data from a control section 109 into a video signal, and outputs it to an adder 114. An adder 114 adds the video signal from the video-signal processing circuit 103, and the character signal from the character signal creation circuit 113, and outputs them to the record regenerative circuit 106 and a switch 117. The record regenerative circuit 106 and the switch 117 are the same as that of the above-mentioned actuation.

[0022] Here, a control section 109 controls the record regenerative circuit 106, a switch 117, and switch 118 grade according to actuation of a control unit 110 of having various actuation keys, such as record, playback, a halt, and mode setting, including a microcomputer.

[0023] Next, concrete actuation of the speech recognition circuit 111 is explained. The sound signal inputted from the sound signal processing circuit 105 is amplified by predetermined gain in pre amplifier 111a, and is outputted to A/D-converter 111b. A/D-converter 111b carries out A/D conversion of the sound signal inputted from pre amplifier 111b, and outputs it to microprocessor 111c. Here, microprocessor 111c is performing the automatic gain control which requires feedback control for pre amplifier 111a so that the sound volume suitable for speech recognition may be inputted.

[0024] To the sound signal from A/D-converter 111b, like the after-mentioned, microprocessor 111c performs speech recognition processing, and outputs the speech recognition actuation situation and speech recognition result to a control section 109 with a data bus. Moreover, microprocessor 111c operates according to the voice mode set up with a switch 117, and mentions later about the actuation. Furthermore, microprocessor 111c outputs the voice pattern obtained by speech recognition processing at the time of voice register mode to memory 111d.

[0025] Drawing 2 is a flow chart which shows the flow of the speech recognition processing which microprocessor 111c performs. Here, the standard pattern set as the object of recognition is created, and the pattern-matching method which performs word speech recognition is used by judging whenever [ input voice and standard pattern's coincidence ].

[0026] In drawing 2, it distinguishes whether speech recognition mode is set up (S201), and when it is in speech recognition mode, speech recognition mode processing is performed. (S202) It distinguishes whether on the other hand, when it is not in speech recognition mode in S201, voice register mode is set up (S203), and when voice register mode is set up, voice register mode processing is performed. (S204)

[0027] Next, concrete actuation in each mode is explained. Drawing 3 is a flow chart which shows actuation of voice register mode processing of drawing 2. If the switch 117 is set to a terminal side, voice register mode processing is started, and a control signal will be outputted to a control section 109 in order to display the list screen of a character title as shown in drawing 5 (a) in S301.

[0028] Here, setting out of a character title is performed in the character setting-out mode set up by the control unit 110. That is, a control section 109 outputs a control signal to the character signal generation circuit 113 so that it may perform the display shown in the monitor 107 at drawing 4, if character setting-out mode is directed by the control unit 110. At this time,

image pick-up actuation by the image pick-up section 102 is not performed, but the background of drawing 4 serves as a predetermined image.

[0029] Like drawing 4, a user is operating a control unit 110, and choosing and determining a desired alphabetic character in 50 sound displays, and inputs a character title. A control section 109 attaches a title number for every title to the character title inputted here, and memorizes it in the character data memory 112. Thus, the item list of the character titles memorized by the character data memory 112 is displayed as a character title list screen of drawing 5 (a) in S301. In addition, the character title list screen of drawing 5 (a) can display the list key of a control unit 110 also by carrying out a depression.

[0030] A display of a character title list screen chooses a character title in S302. Drawing 5 (b) and drawing 5 (c) can show an example of selection of a character title, and the arrow-head cursor on a screen can choose the item of taking up and down and a request according to the arrow key of a control unit 110.

[0031] Thus, selection of which character title performs the input and registration of the voice made to correspond to a character title next. If microprocessor 111c detects whether the switch 118 is turned on in S303 and the switch 118 is turned on in it, the title number of the character title chosen from the control section 109 will be read.

[0032] Next, in S304, detection and the voice analysis of a user's input voice are performed. A user utters a title name here that the selected character title and input voice should be made to correspond, and it should register. For example, if the "nice shot" of the title number 2 is chosen like drawing 5 (b), "\*\*\*\*\* which is not", or utterance which assumed the actual operating condition like "\*\*\*\*\*" will be performed. That is, original voice can be registered by uttering the vocabulary of not only the vocabulary of the character title chosen as a title name but arbitration at the time of registration.

[0033] The uttered voice data determines the start point of a word, and an ending point from an audio amplitude pattern etc., and starts a word while it is changed into the time series of an analysis parameter vector by voice analyses, such as band pass filter analysis.

[0034] Next, a focus extract is performed in S305 and the focus is changed into a feature parameter. Here, the local peak of the obtained spectrum is detected and only these are binary-ization-extracted. Linearity or nonlinear time amount normalization processing is performed to the obtained feature parameter, and a voice pattern is generated. (S306)

[0035] Next, the dependability judging of a voice pattern is performed in S307. Here, it judges whether the generated voice pattern has reached the level which deserves registering as a reference voice pattern. If judged with dependability being inadequate, a voice pattern is not registered, cannot be registered in S308, and it will output a control signal to a control section 109 so that it may display the advice of the purport which needs voice reinput like shown in drawing 6. (S309) It waits for a control section 109 to control the character signal creation circuit 113, and for a switch 118 to be again turned on by the monitor 107 in a predetermined time deed and S303 in this advice display.

[0036] On the other hand, if judged with the dependability of a voice pattern being enough, a control signal will be outputted to a control section 109 in order to display the purport of the completion of registration, while making memory 111d memorize the title number corresponding to a voice pattern and it in S308, and registration actuation will be completed.

[0037] Next, speech recognition mode processing is explained. Drawing 7 is a flow chart which shows actuation of speech recognition mode processing. If the switch 117 is set to b terminal side, it will go into speech recognition mode. If the switch 118 is turned on in S701, while performing detection and the voice analysis of a user's input voice in S702, the signal of the purport which forbids reception of actuation of other actuation keys is outputted to a control section 109. About a feature extraction S703 and the time amount normalization S704, the same actuation as S305 and S306 of above-mentioned drawing 3 is performed.

[0038] Next, in S705, distance count with the input voice pattern vector and reference voice pattern vector of input voice and a reference voice pattern by which matching processing, i.e., time amount normalization, was carried out is performed. It is judged as a word the minimum thing in distance with each reference voice pattern calculated in S705 has been recognized to

be. (S706)

[0039] Next, in S707, the degree of correlation with an input voice pattern and a reference voice pattern, i.e., the dependability judging of whenever [ speech recognition ], is performed. Here, it judges whether the distance of the input voice pattern and reference voice pattern which were calculated is smaller than a predetermined threshold like the above-mentioned. While outputting a control signal to a control section 109 in order to judge whether there being no recognition dependability and to perform a display to that effect, if larger than a threshold, it waits to turn on a switch 118 again in S701. If smaller than a threshold, it will be judged with there being recognition dependability, and in S708, the title number of the recognized voice pattern is read from memory 111d, and it outputs to a control section 109.

[0040] Moreover, if a switch 117 is made into the position c of a termination to perform the title display by speech recognition, speech recognition actuation will be stopped and the display of a character title will not be performed, either.

[0041] Drawing 8 is a flow chart which shows actuation of the control section 109 about a character display. In S801, a control section 109 judges whether the title number was inputted from the speech recognition circuit 111. If it judges that the title number was inputted, the character data corresponding to the title number inputted in S802 will be read from the character data memory 112, and the timer built in while outputting to the character signal creation circuit 113 will be started.

[0042] In S804, when the display of a character title is continued and it exceeds 3 seconds based on the output of a timer while title display time did not exceed 3 seconds, a title display is stopped in S805 and it returns to the usual screen. In addition, although title display time was made into 3 seconds in this gestalt, it is not necessary to be especially 3 seconds, and if it is the time amount a title display is recognized to be, it is good without limit.

[0043] Moreover, drawing 9 and drawing 10 show the example of a display of a character title.

[0044] Thus, in this gestalt, the image which displayed the character title by speech recognition, carried out multiplex [ of the character title ] to the picturized image, and displayed it on it is recordable. Therefore, the key stroke for a title display becomes unnecessary, and a user can express a character as the very easy approach of voice input. Therefore, a display of a character title is attained to desired timing, and the display also of two or more kinds of character titles is attained easily.

[0045] Moreover, the input voice of arbitration can be registered corresponding to a character, and it can indicate by the character by uttering the input voice. Therefore, since a recognition system can be set as the specified speaker when a specified speaker performs voice input, a recognition rate becomes high.

[0046] Furthermore, since the vocabulary of a character and the vocabulary of the input voice to register do not need to be the same, they can display a character in accordance with a user's operating condition.

[0047] Drawing 11 is the block diagram showing the configuration of the digital VTR as second operation gestalt of this invention. The same number is attached about the same component as drawing 1, and the detailed explanation is omitted.

[0048] With VTR of drawing 1, the voice inputted by the microphone 104 has been recognized, and it generated and carried out multiplex [ of the character ] to the picturized video signal. On the other hand, speech recognition is performed to a playback sound signal, and a character title is expressed as this gestalt. That is, as shown in drawing 11, at the time of voice register mode, a switch 115 and a switch 116 are connected to r side edge child, the sound signal from the sound signal processing circuit 105 is inputted into the speech recognition circuit 111, and voice register mode processing is performed according to the flow chart of drawing 3.

[0049] Next, speech recognition mode processing is explained. If the switch 117 is connected to b terminal side, it will go into speech recognition mode. Moreover, at the time of speech recognition mode, it connects with p terminal side, respectively, and a switch 115,116 inputs the regenerative signal from the record regenerative circuit 106 into the speech processing circuit 111.

[0050] Drawing 12 is a flow chart which shows actuation of the speech recognition mode

processing in the second operation gestalt. Generally, since a user utters the voice for character titles in many cases and it is recorded in comparatively loud voice, it is louder than the usual voice level. For this reason, the voice for character titles can be distinguished with the inputted voice level. In S1201, it distinguishes whether voice input level is more than predetermined level in the regenerative signal from the record regenerative circuit 106. With [ the inputted voice level ] predetermined [ more than ], voice detection and a voice analysis are performed in S1202, and the following actuation of S1203-S1208 is the same as actuation of drawing 7 of S702-S707.

[0051] Thus, in this gestalt, multiplex [ of the character title ] can be carried out to the image reproduced by speech recognition, and it can be displayed on it. Therefore, the key stroke for a title display becomes unnecessary, and a user can express a character as the very easy approach of voice input. Therefore, a display of a character title is attained to desired timing, and the display also of two or more kinds of character titles is attained easily.

[0052] Moreover, the input voice of arbitration can be registered corresponding to a character, and it can indicate by the character by uttering the input voice. Therefore, since a recognition system can be set as the specified speaker when a specified speaker performs voice input, a recognition rate becomes high.

[0053] Furthermore, since the vocabulary of a character and the vocabulary of the input voice to register do not need to be the same, they can display a character in accordance with a user's operating condition.

[0054] In addition, although this gestalt explained the case where this invention was applied to video camera one apparatus VTR, besides this, this invention can be applied to a record regenerative apparatus with a character display function, and it has the same effectiveness.

[0055] Moreover, although the alphabetic character character was displayed, you may make it express a graphic form etc. as the above-mentioned operation gestalt as a character besides an alphabetic character. Furthermore, the display position of a character can also be displayed on the location of not only the location shown in drawing 9 and drawing 10 but arbitration, and can be changed also about the method of presentation.

[0056]

[Effect of the Invention] Since it is the configuration in which the character title display with voice is possible according to this invention as explained above, two or more character displays are enabled easily, and a character title can be expressed on a screen as more effective timing.

[0057] Moreover, since it has the composition of registering input voice corresponding to a character, the rate of speech recognition of a specified speaker becomes high.

[0058] Furthermore, since it has the composition that the vocabulary of a character may differ from the vocabulary of the input voice to register, a character can be displayed in accordance with a user's operating condition.

---

[Translation done.]

**\* NOTICES \***

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

**DESCRIPTION OF DRAWINGS****[Brief Description of the Drawings]**

**[Drawing 1]** It is drawing showing the configuration of video camera one apparatus VTR to which this invention is applied as first operation gestalt.

**[Drawing 2]** It is the flow chart of the speech recognition processing which the microprocessor of drawing 1 performs.

**[Drawing 3]** It is the flow chart which shows actuation of voice register mode processing of drawing 2.

**[Drawing 4]** It is drawing showing the example of a monitor display at the time of character setting-out mode.

**[Drawing 5]** It is drawing showing the example of a display of a character title list screen.

**[Drawing 6]** It is drawing showing the example of a display of the voice reinput advice display at the time of being judged with it being inadequate in dependability.

**[Drawing 7]** It is the flow chart which shows actuation of speech recognition mode processing of drawing 2.

**[Drawing 8]** It is the flow chart which shows the actuation about the character display which the control section of drawing 1 performs.

**[Drawing 9]** It is drawing showing the example of a display of a character title.

**[Drawing 10]** It is drawing showing the example of a display of a character title.

**[Drawing 11]** It is drawing showing the configuration of video camera one apparatus VTR to which this invention is applied as second operation gestalt.

**[Drawing 12]** It is the flow chart which shows actuation of the speech recognition mode processing in the second operation gestalt.

**[Translation done.]**

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-101919

(P2000-101919A)

(43)公開日 平成12年4月7日(2000.4.7)

(51)Int.Cl.'

H 04 N  
5/278  
5/225

識別記号

F I

H 04 N  
5/278  
5/225

マーク(参考)

5 C 0 2 2  
5 C 0 2 3

審査請求 未請求 請求項の数27 O L (全 12 頁)

(21)出願番号

特願平10-273129

(22)出願日

平成10年9月28日(1998.9.28)

(71)出願人

000001007  
キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72)発明者

竹井 浩文  
東京都大田区下丸子3丁目30番2号キヤノン株式会社内

(74)代理人

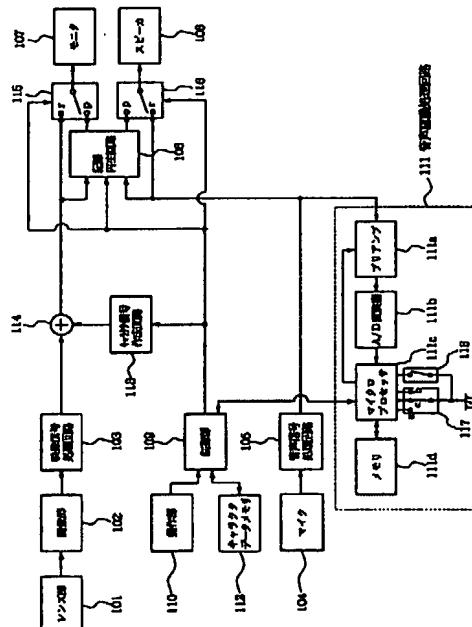
100069877  
弁理士 丸島 優一  
Fターム(参考) 5C022 AA11 AC42 AC72 AC79  
5C023 AA18 AA37 BA11 CA01 CA06  
DA08

(54)【発明の名称】 信号処理装置及び方法、再生装置及び方法、及び撮像装置

(57)【要約】

【課題】 容易に複数のキャラクタ表示を可能とし、より効果的なタイミングでキャラクタタイトルを画面に表示可能とする。

【解決手段】 信号処理装置は、音声信号と映像信号を処理する装置であって、音声信号を入力する入力手段と、所定のキャラクタに対応させて前記入力音声を登録する音声登録モードと前記入力音声中の所定の音声信号を認識する音声認識モードとを有し、前記入力された音声信号を認識する音声認識手段と、前記認識された音声信号に対応するキャラクタ信号を発生する発生手段と、前記発生手段により発生されたキャラクタ信号に対応したキャラクタを表示する表示手段と、前記発生手段により発生されたキャラクタ信号を映像信号と共に記録する記録手段とを備えて構成されている。



1

## 【特許請求の範囲】

【請求項 1】 音声信号と映像信号を処理する装置であつて、音声信号を入力する入力手段と、所定のキャラクタに対応させて前記入力音声を登録する音声登録モードと前記入力音声中の所定の音声信号を認識する音声認識モードとを有し、前記入力された音声信号を認識する音声認識手段と、前記認識された音声信号に対応するキャラクタ信号を発生する発生手段と、前記発生手段により発生されたキャラクタ信号に対応したキャラクタを表示する表示手段と、前記発生手段により発生されたキャラクタ信号を映像信号と共に記録する記録手段とを備えることを特徴とする信号処理装置。

【請求項 2】 前記音声認識手段は、前記入力音声中より認識することを特徴とする請求項 1 記載の信号処理装置。

【請求項 3】 前記音声認識手段は、キャラクタに対応させて前記入力音声を登録する登録モードを備えることを特徴とする請求項 1 もしくは 2 記載の信号処理装置。

【請求項 4】 前記音声認識手段は、メモリを有し、前記メモリに対して前記キャラクタ信号に係る制御情報と前記入力音声とを対応づけて記憶することを特徴とする請求項 3 記載の信号処理装置。

【請求項 5】 前記登録モードにおいて、前記音声認識手段は、前記表示手段によりキャラクタ表示すると共に、前記キャラクタに対応づけて前記入力音声を登録することを特徴とする請求項 3 もしくは 4 記載の信号処理装置。

【請求項 6】 前記登録モードにおいて、前記音声認識手段は、前記表示手段により複数の前記キャラクタを表示すると共に、複数の前記キャラクタに対応づけて前記入力音声を登録することを特徴とする請求項 5 記載の信号処理装置。

【請求項 7】 前記登録モード時において前記表示手段により表示された複数のキャラクタを選択する選択手段を備え、前記音声認識手段は、前記選択手段により選択されたキャラクタに対して入力された前記入力音声を登録することを特徴とする請求項 6 記載の信号処理装置。

【請求項 8】 前記音声認識手段は、前記登録モードにおいて、前記入力音声信号が登録に適しているか否かを判定し、当該判定結果に応じて前記入力音声を登録することを特徴とする請求項 3～7 のいずれかに記載の信号処理装置。

【請求項 9】 前記音声認識手段は、前記入力音声信号が登録に不適切であると判定した場合には、当該入力音声信号の登録を禁止すると共に、所定の表示を行うべく前記キャラクタ発生手段と前記表示手段とを制御することを特徴とする請求項 8 記載の信号処理装置。

【請求項 10】 前記音声認識手段は、前記認識された音声信号に基づいて前記キャラクタに係る制御情報を前記発生手段に出力することを特徴とする請求項 1～9 の

2

## いずれかに記載の信号処理装置。

【請求項 11】 前記音声認識手段は、前記制御情報を記憶するメモリを有し、前記認識された音声信号に基づいて前記制御情報を前記メモリより読み出して前記発生手段に出力する請求項 10 記載の信号処理装置。

【請求項 12】 前記メモリは更に、前記制御情報と音声情報を対応づけて記憶し、前記音声認識手段は、前記音声情報を用いて前記入力音声信号を認識すると共に、前記認識された入力音声信号に対応した制御情報を前記メモリより読み出して前記発生手段に出力することを特徴とする請求項 11 記載の信号処理装置。

【請求項 13】 前記音声認識手段は、前記メモリに記憶されている音声情報と前記入力音声信号との相関の度合いを検出し、当該検出結果に基づいて前記入力音声信号を認識することを特徴とする請求項 12 記載の信号処理装置。

【請求項 14】 前記音声認識手段は、前記相関の度合いが低い場合には、前記メモリからの前記制御情報を前記発生手段に出力することを禁止すると共に、所定の表示を行うべく前記キャラクタ発生手段と前記表示手段を制御することを特徴とする請求項 13 記載の信号処理装置。

【請求項 15】 前記発生手段は、複数のキャラクタにそれぞれ対応したキャラクタデータを記憶する記憶手段を備え、前記音声認識手段より出力された制御情報に応じたキャラクタデータを前記記憶手段より読み出すことを特徴とする請求項 10～14 のいずれかに記載の信号処理装置。

【請求項 16】 被写体を撮像し、映像信号を出力する撮像手段を備え、前記表示手段は前記発生手段からのキャラクタ信号に対応したキャラクタを前記撮像手段から出力された映像信号に対応した映像に多重して表示することを特徴とする請求項 1～15 のいずれかに記載の信号処理装置。

【請求項 17】 前記音声認識手段は、前記音声登録モードと前記音声認識モードとを設定するマニュアル設定手段を備えることを特徴とする請求項 1 記載の信号処理装置。

【請求項 18】 前記音声登録モード及び前記音声認識モードにおいて、登録動作及び認識動作を始める為のトリガースイッチとなる音声入力スイッチを備えることを特徴とする請求項 1 もしくは 17 記載の信号処理装置。

【請求項 19】 前記発生手段は、前記キャラクタ信号の発生を開始してから所定の時間後に前記キャラクタ信号の発生を停止することを特徴とする請求項 1～18 のいずれかに記載の信号処理装置。

【請求項 20】 音声信号と映像信号を再生する装置であつて、音声信号を再生する再生手段と、前記再生された音声信号を認識する音声認識手段と、前記認識された音声信号に対応するキャラクタ信号を発生する発生手段

50

と、前記発生手段により発生されたキャラクタ信号に対応したキャラクタを表示する表示手段とを備える再生装置。

【請求項21】 前記音声認識手段は、所定のキャラクタに対応させて入力音声を登録する音声登録モードと、前記再生音声中の所定の音声信号を認識する音声認識モードとを有することを特徴とする請求項20記載の再生装置。

【請求項22】 前記音声認識手段は、前記音声登録モードと前記音声認識モードとを設定するマニュアル設定手段を備えることを特徴とする請求項21記載の信号処理装置。

【請求項23】 前記音声登録モードにおいて、認識動作を始める為のトリガースイッチとなる音声入力スイッチを備えることを特徴とする請求項21もしくは22記載の再生装置。

【請求項24】 前記音声認識手段は、再生音声信号レベルが所定のレベル以上であるか否かの判定を行い、当該判定結果に従って認識動作を行うことを特徴とする請求項20～23のいずれかに記載の再生装置。

【請求項25】 前記音声認識手段は、前記再生音声レベルが所定のレベル以上であると判定した場合に、認識動作を開始することを特徴とする請求項24記載の再生装置。

【請求項26】 音声信号を入力し、キャラクタに対応させて前記入力音声信号を登録し、前記登録された音声信号に基づいて前記入力された音声信号を認識し、前記認識された音声信号に対応するキャラクタ信号を発生し、前記発生手段により発生されたキャラクタ信号に対応するキャラクタを表示すると共に、前記発生手段により発生されたキャラクタ信号を映像信号と共に記録することを特徴とする信号処理方法。

【請求項27】 音声信号を再生し、前記再生された音声信号を認識し、前記認識された音声信号に対応するキャラクタ信号を発生し、前記発生手段により発生されたキャラクタ信号に対応したキャラクタを表示することを特徴とする再生方法。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】 本発明は信号処理装置及び方法、再生装置及び方法、及び撮像装置に関し、特には、キャラクタ表示を伴う装置及び方法に関するものである。

##### 【0002】

【従来の技術】 従来より、ビデオカメラで撮影したビデオ信号を処理する装置としてビデオカメラ一体型VTRが知られている。そして、ビデオカメラ一体型VTRでは、キャラクタ表示回路を備え、撮影した映像信号にキャラクタタイトルを合成して表示する機能を備えたものがあった。

【0003】 この種の装置において、画面にキャラクタタイトルを表示させる場合、撮影者がスイッチによる操作を行っていた。

##### 【0004】

【発明が解決しようとする課題】 しかし、このような方法では、操作が煩雑であるだけでなく、手動による操作のため、キャラクタタイトルの表示タイミングが画面の動きと一致しない。そのため、画面上にタイミングよくキャラクタタイトルを表示させることは困難であった。

10 特に複数のキャラクタタイトルを画面内容に応じて表示させたい場合などは一層困難であった。

【0005】 本発明は前述の如き問題を解決することを目的とする。

【0006】 本発明の更に他の目的は、容易に複数のキャラクタ表示を可能とし、より効果的なタイミングでキャラクタタイトルを画面に表示可能とするところにある。

##### 【0007】

【課題を解決するための手段】 このような目的下において、本件請求項1に記載の発明による信号処理装置においては、音声信号を入力する入力手段と、所定のキャラクタに対応させて前記入力音声を登録する音声登録モードと前記入力音声中の所定の音声信号を認識する音声認識モードとを有し、前記入力された音声信号を認識する音声認識手段と、前記認識された音声信号に対応するキャラクタ信号を画面上に表示する表示手段と、前記表示されたキャラクタ信号を映像信号と共に記録する記録手段とを備える構成とした。

##### 【0008】

20 【発明の実施の形態】 以下、図面を用いて本発明の実施形態について詳細に説明する。

【0009】 図1は本発明が適用されるカメラ一体型デジタルVTRの構成を第一の実施形態として示すブロック図である。

【0010】 まず、通常の記録再生時の動作について説明する。図において、レンズ部101はレンズ、ズーム機構、アイリス等を含み、被写体光を撮像部102に導く。撮像部102はCCDを使って被写体を撮像し、得られた被写体光を電気信号に変換し、映像信号処理回路40103に出力する。

【0011】 映像信号処理回路103は撮像部102より出力された信号に対してクランプ、ア補正、ホワイトバランス等、撮像された信号をテレビジョン信号に準拠した映像信号とするための周知の処理を施し、加算器114を通して記録再生回路106及びスイッチ115に出力する。ここで、加算器114は後述の如くキャラクタ信号作成回路113より出力されたキャラクタ信号を加算器114を映像信号処理回路103より出力された映像信号に加算し、記録再生回路106及びスイッチ115に出力する。

50

【0012】記録再生回路106は映像信号処理回路103より出力された映像信号に対してA/D変換、DC-T、ハフマン符号化等周知の可変長符号化を用いて符号化することでの情報量を圧縮し、圧縮した映像信号に対して同期、ID等の付加、誤り訂正符号化等の記録する為に必要な処理を施し、記録に適した形態に変更する。更に、この様な処理を施した映像信号を磁気テープ上に記録する。スイッチ115は映像信号記録時にr側端子に接続され、映像信号処理回路103より出力された映像信号をモニタ107に出力する。

【0013】この様に記録された映像信号は記録再生回路106により再生され、誤り訂正、可変長復号、逆DC-T等の処理を施された後、スイッチ115に出力される。スイッチ115は映像信号再生時にp側端子に接続され、記録再生回路106からの再生信号をモニタ107に出力する。

【0014】一方、マイク104は音声を集音し、得られた音声を音声信号処理回路105に出力する。音声信号処理回路105はマイクからの音声信号に対してレベル調整、ノイズリダクション等の周知の処理を施し、記録再生回路106、スイッチ116及び音声認識回路111に出力する。記録再生回路106は音声信号処理回路105からの音声信号に対してA/D変換、シャフリング、誤り訂正等の周知の処理を施し、磁気テープ上に記録する。スイッチ116は音声信号記録時にr側端子に接続され、音声信号処理回路105より出力された音声信号をスピーカ108に出力する。音声認識回路111については後述する。

【0015】この様に記録された音声信号は記録再生回路106により再生され、誤り訂正、デシャフリング等の処理を施された後、スイッチ116に出力される。スイッチ116は音声信号再生時にp側端子に接続され、記録再生回路106からの再生信号をスピーカ108に出力する。

【0016】次に本形態における音声認識回路111の動作について説明する。一般的に音声認識装置には、話者を限定する特定話者用と話者を限定せず誰の音声でも認識する不特定話者用がある。特定話者用は認識する語彙を予め発声させ、登録しておき、この登録したデータに基づいて認識を行うものである。この特定話者用の装置によれば、使用する特定の話者に認識系を設定できる為、システムの負荷が軽くなると共に高い認識率が期待でき、言語にも依存されにくい。

【0017】一方、不特定話者用は登録動作が不要なため、話者を選び、すぐに音声認識を動作させる簡便性はあるが、認識精度を上げる為には演算装置、メモリとも大規模なシステムが必要となる。本形態の如きカメラ一体型VTRの場合、ユーザはほぼ一個人に限定されることと小型かつ低コストを絶対条件とすること等から、特定話者であり、更に特定語彙を対象とする音声認識装

置が適しているといえる。

【0018】この様な背景から本形態においては、特定話者に対応した音声認識システムを用いている。このため、認識する語彙をユーザに予め発声させ、その音声データを参照音声パターンとして登録する音声登録モードと、入力音声と参照音声パターンとのマッチングを行う音声認識モードとを備えている。

【0019】図1において音声モードの切換はスイッチ117で行い、スイッチ117は音声入力機能をOFFするポジションc、音声登録モードa及び音声認識モードbの3ポジションを選択可能である。スイッチ118は登録、認識モード時にユーザが発声する音声を入力する際のトリガースイッチとなる音声入力スイッチである。

【0020】次に、音声認識結果によりキャラクタ信号を発生し、モニタ107に表示する場合について述べる。マイク104から音声信号処理回路105を介して音声認識回路111に入力された音声信号は、後述の如く音声認識回路111において音声認識処理を施される、音声認識回路111はその音声認識結果を制御部109に出力する。ここで音声認識回路111は、スイッチ117により音声登録モード若しくは音声認識モードに設定されている時の動作する。

【0021】制御部109は音声認識回路111からの音声認識結果に応じたキャラクタデータをキャラクタデータメモリ112より読み込み、キャラクタ信号作成回路113に出力する。キャラクタ信号作成回路113は制御部109からのキャラクタデータを映像信号に変換し、加算器114に出力する。加算器114は、映像信号処理回路103からの映像信号とキャラクタ信号作成回路113からのキャラクタ信号を加算し、記録再生回路106、スイッチ117に出力する。記録再生回路106及びスイッチ117は前述の動作と同様である。

【0022】ここで、制御部109はマイクロコンピュータを含み、記録、再生、停止、モード設定等各種操作キーを有する操作部110の操作に応じて記録再生回路106、スイッチ117、スイッチ118等を制御する。

【0023】次に、音声認識回路111の具体的な動作について説明する。音声信号処理回路105より入力された音声信号は、ブリアンプ111aにおいて所定ゲインで増幅され、A/D変換器111bに出力される。A/D変換器111bはブリアンプ111bより入力された音声信号をA/D変換し、マイクロプロセッサ111cに出力する。ここで、マイクロプロセッサ111cは音声認識に適した音量が入力されるようにブリアンプ111aにフィードバック制御がかかるオートゲインコントロールを行っている。

【0024】マイクロプロセッサ111cはA/D変換器111bからの音声信号に対して後述の如く音声認識処理を行い、その音声認識動作状況と音声認識結果とをデ

ータバスによって制御部109に出力する。また、マイクロプロセッサ111cはスイッチ117で設定される音声モードに応じて動作し、その動作については後述する。更に、マイクロプロセッサ111cは音声登録モード時に音声認識処理で得られた音声パターンをメモリ11dに出力する。

【0025】図2はマイクロプロセッサ111cが行う音声認識処理の流れを示すフローチャートである。ここでは、認識の対象となる標準パターンを作成し、入力音声と標準パターンの一一致度を判定することにより単語音声認識を行うパターンマッチング方式を用いている。

【0026】図2において、音声認識モードが設定されているか否かを判別し(S201)、音声認識モードであった場合には音声認識モード処理を実行する。(S202)一方、S201で音声認識モードでなかった場合は、音声登録モードが設定されているか否かを判別し(S203)、音声登録モードが設定されていた場合には音声登録モード処理を実行する。(S204)

【0027】次に各モードの具体的な動作について説明する。図3は図2の音声登録モード処理の動作を示すフローチャートである。スイッチ117がa端子側に設定されていると、音声登録モード処理をスタートし、S301において図5(a)に示す様なキャラクタタイトルの一覧画面を表示するべく制御部109に制御信号を出力する。

【0028】ここで、キャラクタタイトルの設定は操作部110により設定されるキャラクタ設定モードにおいて行う。即ち、制御部109は操作部110によりキャラクタ設定モードが指示されるとモニタ107に図4に示した表示を行なうべくキャラクタ信号生成回路113に制御信号を出力する。この時、撮像部102による撮像動作は行われておらず、図4の背景は所定の映像となる。

【0029】ユーザは図4の様に50音表示において、操作部110を操作して所望の文字を選択、決定することで、キャラクタタイトルを入力する。制御部109は、ここで入力したキャラクタタイトルに対してタイトルごとにタイトル番号を付け、キャラクタデータメモリ112に記憶する。この様にキャラクタデータメモリ112に記憶されたキャラクタタイトルの項目一覧がS301において図5(a)のキャラクタタイトル一覧画面として表示される。なお、図5(a)のキャラクタタイトル一覧画面は操作部110の一覧キーを押下することでも表示可能である。

【0030】キャラクタタイトル一覧画面が表示されると、S302においてキャラクタタイトルの選択を行う。図5(b)及び図5(c)はキャラクタタイトルの選択の一例を示すものであり、画面上の矢印カーソルが操作部110の矢印キーに応じて上下し、所望の項目を選択することができる。

【0031】この様に何れかのキャラクタタイトルを選択すると、次にキャラクタタイトルに対応させる音声の入力及び登録を行う。S303においてスイッチ118がONになっているかをマイクロプロセッサー111cが検知し、スイッチ118がONになっていれば制御部109から選択されたキャラクタタイトルのタイトル番号を読み込む。

【0032】次にS304においてユーザの入力音声の検出及び音声分析を行う。ここでユーザは選択されたキャラクタタイトルと入力音声とを対応させて登録すべく、タイトル名を発声する。例えば、図5(b)のようにタイトル番号2の「ナイスショット！」が選択されていれば、「ないすしょっと！」若しくは「ふぁいん！」のように実際の使用状況を想定した発声を行う。即ち、タイトル名として選択されたキャラクタタイトルの語彙だけでなく、任意の語彙を登録時に発声することで独自の音声を登録することができる。

【0033】発声した音声データはバンドパスフィルタ分析等の音声分析により分析パラメータベクトルの時系列に変換されると共に、音声の振幅パターン等から単語の開始点、終了点を決定し単語の切り出しを行う。

【0034】次にS305において特徴点抽出を行い、特徴点を特徴パラメータに変換する。ここでは得られたスペクトルのローカルピークを検出し、これらのみを2値化抽出する。得られた特徴パラメータに対して線形又は非線形の時間正規化処理を行い、音声パターンを生成する。(S306)

【0035】次にS307において音声パターンの信頼性判定が行われる。ここでは、生成された音声パターンが参照音声パターンとして登録するのに値するレベルに達しているかを判定する。信頼性が不十分であると判定されると、音声パターンは登録されず、S308において登録不可能であり、図6に示した様に、音声再入力を必要とする旨の勧告を表示を行なうべく制御部109に制御信号を出力する。(S309)制御部109はキャラクタ信号作成回路113を制御して、モニタ107にこの勧告表示を所定時間行い、S303において再度スイッチ118がONされるのを待つ。

【0036】一方、音声パターンの信頼性が十分であると判定されると、S308において音声パターンとそれに対応するタイトル番号をメモリ111dに記憶せると共に登録完了の旨の表示を行なうべく制御部109に制御信号を出力し、登録動作が完了する。

【0037】次に音声認識モード処理について説明する。図7は音声認識モード処理の動作を示すフローチャートである。スイッチ117がb端子側に設定されていると、音声認識モードに入る。S701においてスイッチ118がONになっていればS702においてユーザの入力音声の検出及び音声分析を行うと共に、他の操作キーの動作の受付を禁止する旨の信号を制御部109に

出力する。特徴抽出S703、時間正規化S704については、前述の図3のS305、S306と同様の動作を行う。

【0038】次にS705において入力音声と参照音声パターンとのマッチング処理、即ち、時間正規化された入力音声パターンベクトルと参照音声パターンベクトルとの距離計算を行う。S705において計算された各参照音声パターンとの距離の中で最小のものが認識された単語として判定される。(S706)

【0039】次にS707において、入力音声パターンと参照音声パターンとの相関の度合い、即ち音声認識度の信頼性判定が行われる。ここでは、前述の如く計算された入力音声パターンと参照音声パターンとの距離が所定の閾値よりも小さいかを判定する。閾値よりも大きければ認識信頼性がないと判定し、その旨の表示を行なうべく制御部109に制御信号を出力すると共に、S701において再度スイッチ118がONされるのを待つ。閾値よりも小さければ認識信頼性があると判定され、S708において、認識された音声パターンのタイトル番号をメモリ111dより読み出し、制御部109に出力する。

【0040】また、音声認識によるタイトル表示を行いたくない場合は、スイッチ117を中止のポジションcにすれば、音声認識動作が中止され、キャラクタタイトルの表示も行われない。

【0041】図8はキャラクタ表示に関する制御部109の動作を示すフローチャートである。S801において、制御部109は音声認識回路111からタイトル番号が入力されたか否かを判断する。タイトル番号が入力されたと判断すると、S802において入力されたタイトル番号に対応するキャラクタデータをキャラクタデータメモリ112より読み込み、キャラクタ信号作成回路113に出力すると共に内蔵するタイマをスタートする。

【0042】S804においては、タイマの出力に基づき、タイトル表示時間が3秒を超えない間はキャラクタタイトルの表示を継続し、3秒を超えた時点でS805においてタイトル表示を中止して通常の画面に戻す。なお、本形態においてはタイトル表示時間を3秒としたが、特に3秒である必要はなく、タイトル表示が認識される時間であればいくらでもよい。

【0043】また、図9及び図10はキャラクタタイトルの表示例を示したものである。

【0044】このように、本形態においては、音声認識によりキャラクタタイトルを表示し、撮像した映像にキャラクタタイトルを多重して表示した映像を記録することができます。従って、ユーザはタイトル表示のためのキー操作が不要となり、音声入力というきわめて簡単な方法でキャラクタを表示できる。そのため、所望のタイミングでキャラクタタイトルを表示可能となり、又、複数

種類のキャラクタタイトルも容易に表示可能となる。

【0045】また、キャラクタに対応して任意の入力音声を登録し、その入力音声を発声することで、キャラクタ表示することができる。従って、特定話者が音声入力を用いる場合には、その特定話者に認識系を設定できるため、認識率が高くなる。

【0046】更に、キャラクタの語彙と登録する入力音声の語彙は同一である必要がないため、ユーザの使用状況にあわせてキャラクタを表示可能である。

【0047】図11は本発明の第二の実施形態としてのデジタルVTRの構成を示すブロック図である。図1と同様の構成要素については同一番号を付し、その詳細な説明は省略する。

【0048】図1のVTRでは、マイク104により入力された音声を認識してキャラクタを発生し、撮像された映像信号に多重した。これに対し、本形態では、再生音声信号に対して音声認識を行い、キャラクタタイトルの表示を行う。即ち、図11に示すように音声登録モード時には、スイッチ115、スイッチ116をr側端子に接続して音声信号処理回路105からの音声信号を音声認識回路111に入力し、図3のフローチャートに従って音声登録モード処理を実行する。

【0049】次に音声認識モード処理について説明する。スイッチ117がb端子側に接続されていると、音声認識モードに入る。また、音声認識モード時には、スイッチ115、116はそれぞれr端子側に接続され、記録再生回路106からの再生信号を音声処理回路111に入力する。

【0050】図12は第二の実施形態における音声認識モード処理の動作を示すフローチャートである。一般に、キャラクタタイトル用音声は、ユーザが発声することが多く、比較的大きい声で録音される為、通常の音声レベルよりも大きい。このため、入力された音声レベルによって、キャラクタタイトル用音声を判別する事ができる。S1201において記録再生回路106からの再生信号において、音声入力レベルが所定レベル以上であるか否かを判別する。入力された音声レベルが所定以上であれば、S1202において音声検出、音声分析を行い、以下のS1203～S1208の動作は図7のS702～S707の動作と同様である。

【0051】このように、本形態においては、音声認識により再生された映像にキャラクタタイトルを多重して表示することができる。従って、ユーザはタイトル表示のためのキー操作が不要となり、音声入力というきわめて簡単な方法でキャラクタを表示できる。そのため、所望のタイミングでキャラクタタイトルを表示可能となり、又、複数種類のキャラクタタイトルも容易に表示可能となる。

【0052】また、キャラクタに対応して任意の入力音声を登録し、その入力音声を発声することで、キャラク

11

タ表示することができる。従って、特定話者が音声入力を行う場合には、その特定話者に認識系を設定できるため、認識率が高くなる。

【0053】更に、キャラクタの語彙と登録する入力音声の語彙は同一である必要がないため、ユーザの使用状況にあわせてキャラクタを表示可能である。

【0054】なほ、本形態ではビデオカメラ一体型VTRに対して本発明を適用した場合について説明したが、これ以外にも、キャラクタ表示機能のある記録再生装置に対して本発明を適用可能であり、同様の効果を有する。

【0055】また、前述の実施形態では、文字キャラクタを表示したが、文字以外にも例えば图形等をキャラクタとして表示する様にしても良い。更に、キャラクタの表示位置も図9、図10に示した位置だけでなく、任意の位置に表示可能であり、表示方法についても変更可能である。

【0056】

【発明の効果】以上説明したように、本発明によれば、音声によるキャラクタタイトル表示が可能な構成であるので、容易に複数のキャラクタ表示を可能とし、より効果的なタイミングでキャラクタタイトルを画面に表示することができる。

【0057】また、キャラクタに対応して入力音声を登録する構成となっているので、特定話者の音声認識率は高くなる。

【0058】更に、キャラクタの語彙と登録する入力音声の語彙が異なっていてもよい構成となっているので、\*

12

\* ユーザの使用状況にあわせてキャラクタを表示可能である。

【図面の簡単な説明】

【図1】本発明が適用されるビデオカメラ一体型VTRの構成を第一の実施形態として示す図である。

【図2】図1のマイクロプロセッサが行う音声認識処理のフローチャートである。

【図3】図2の音声登録モード処理の動作を示すフローチャートである。

【図4】キャラクタ設定モード時のモニタ表示例を示す図である。

【図5】キャラクタタイトル一覧画面の表示例を示す図である。

【図6】信頼性不十分と判定された場合の音声再入力警告表示の表示例を示す図である。

【図7】図2の音声認識モード処理の動作を示すフローチャートである。

【図8】図1の制御部が行うキャラクタ表示に関する動作を示すフローチャートである。

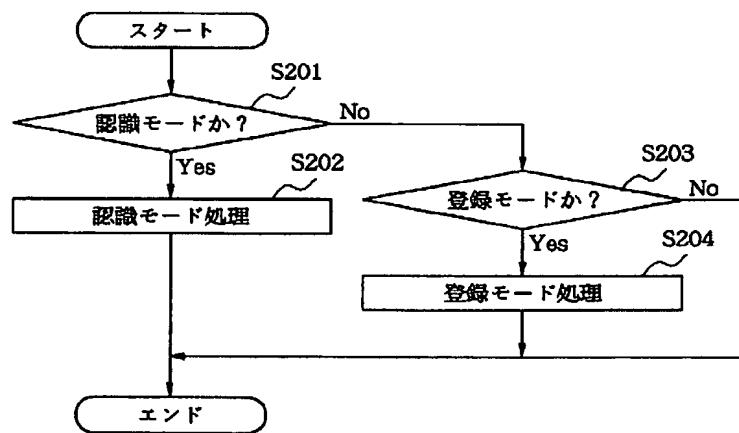
【図9】キャラクタタイトルの表示例を示す図である。

【図10】キャラクタタイトルの表示例を示す図である。

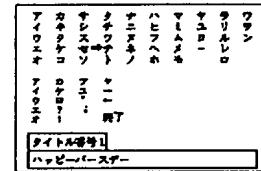
【図11】本発明が適用されるビデオカメラ一体型VTRの構成を第二の実施形態として示す図である。

【図12】第二の実施形態における音声認識モード処理の動作を示すフローチャートである。

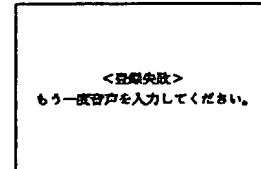
【図2】



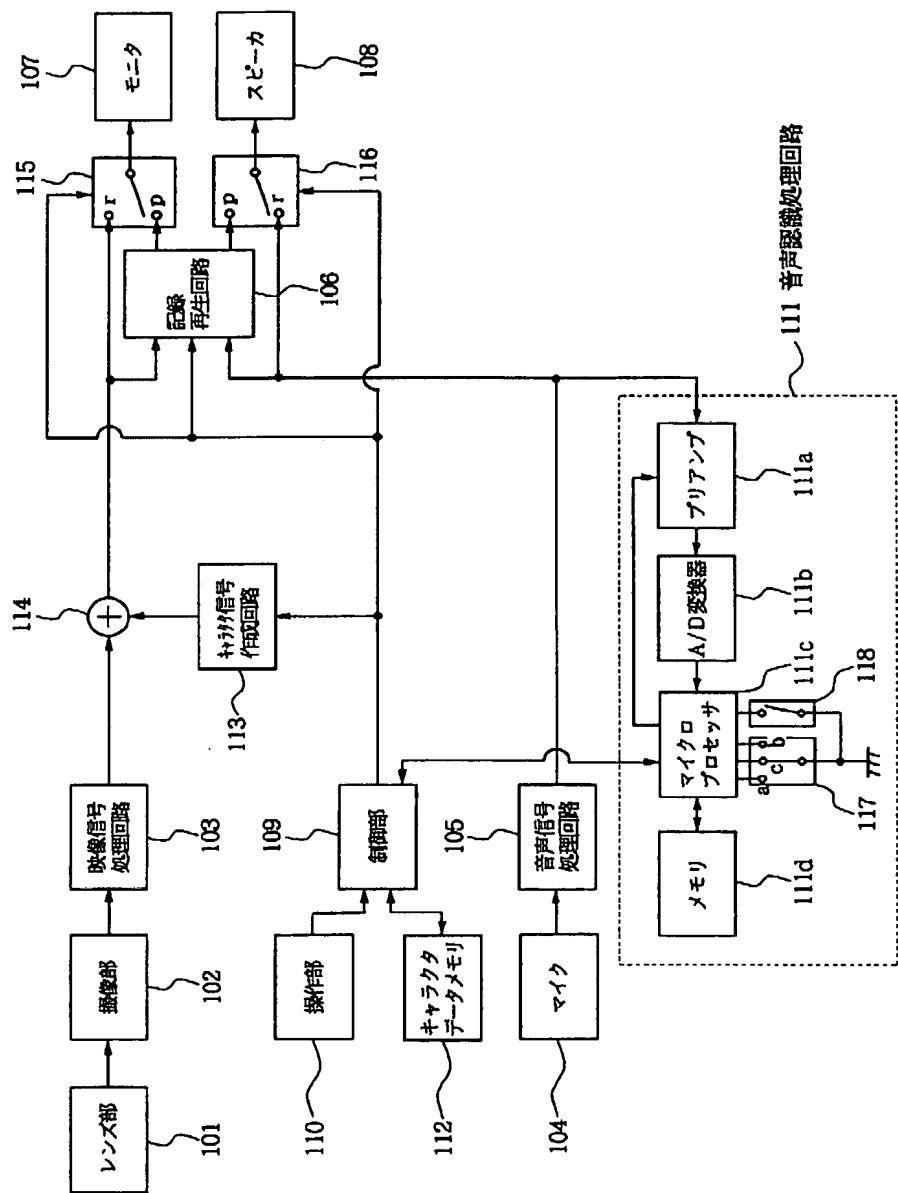
【図4】



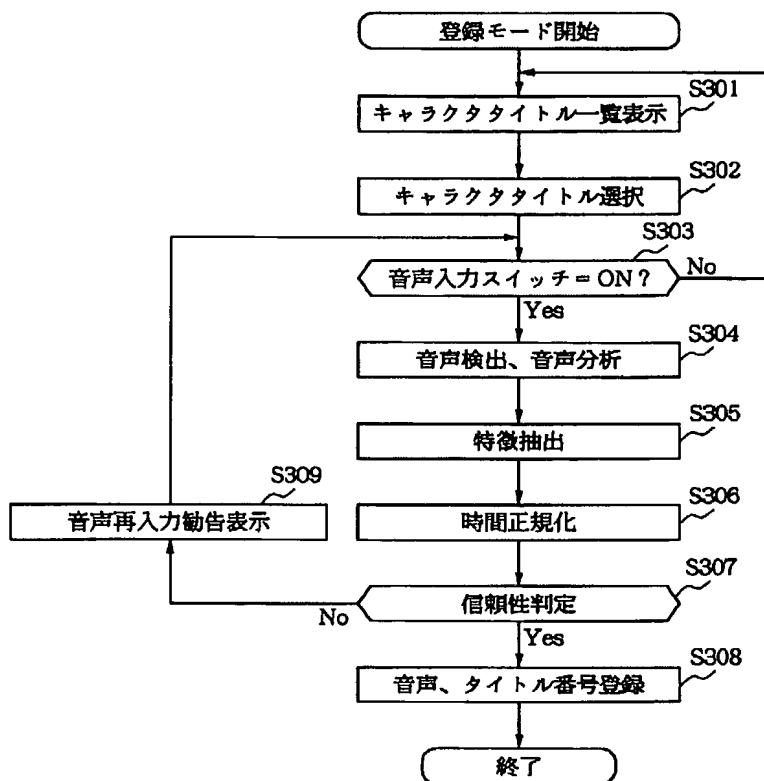
【図6】



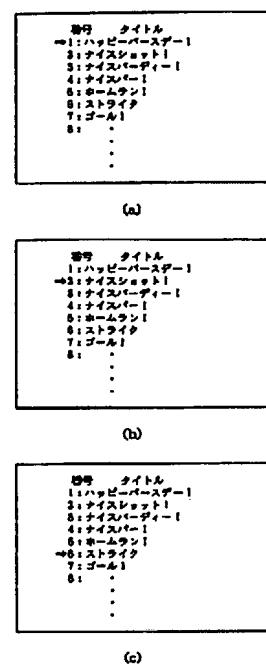
【図1】



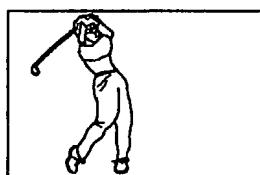
【図3】



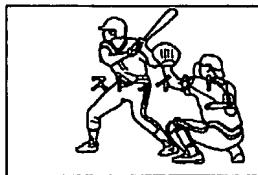
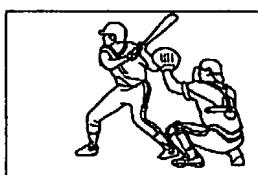
【図5】



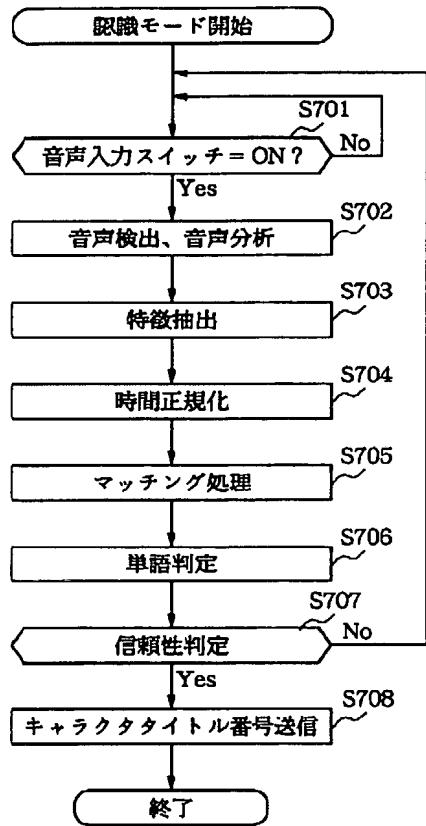
【図9】



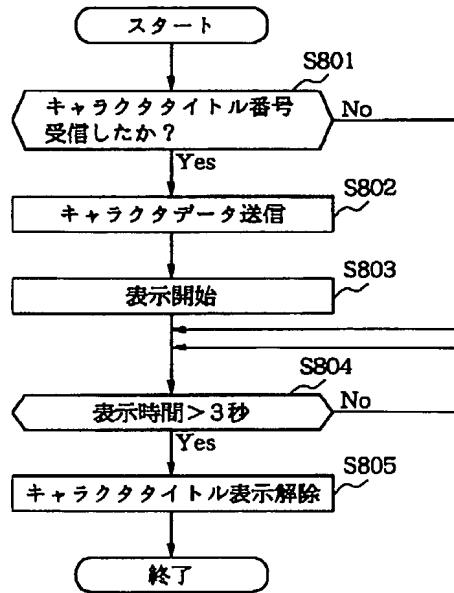
【図10】



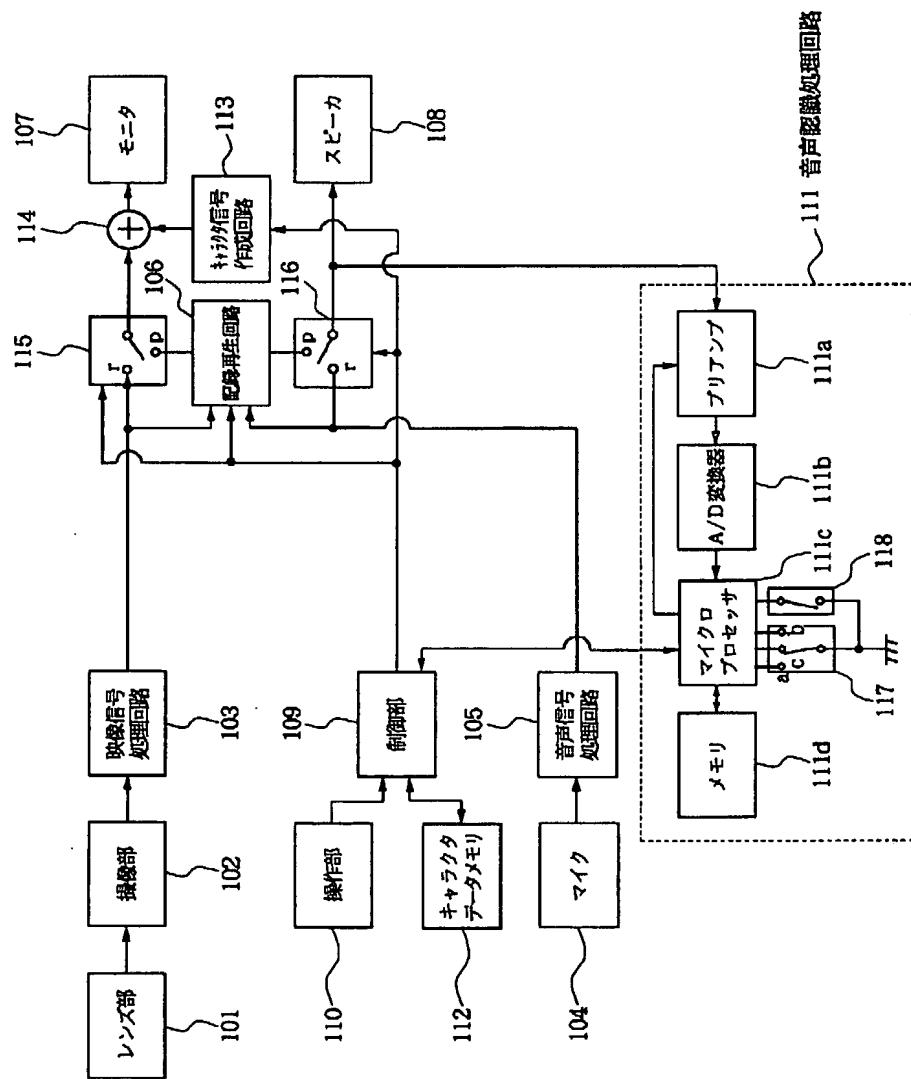
【図7】



【図8】



【図11】



【図12】

